Title: Molecular patterns of cancer colonisation in lymph nodes of breast cancer patients

Running title: Transcriptional patterns of pre- and metastatic lymph nodes in breast cancers

List of authors: Gaurav Chatterjee[1,2,3,*], Trupti Pai [1,2,3,*], Thomas Hardiman[1,2], Kelly Avery-Kiejda[4], Rodney J Scott[4], Jo Spencer[5], Sarah E Pinder[2], Anita Grigoriadis[1,2,6$]

* contributed equally

Full affiliations of all authors

[1] Cancer Bioinformatics, King's College London, Innovation Hub, Cancer Centre at Guy's Hospital, Great Maze Pond, London SE1 9RT, UK.

[2] School of Cancer & Pharmaceutical Sciences, CRUK King's Health Partners Centre, King's College London, Innovation Hub, Comprehensive Cancer Centre at Guy's Hospital, Great Maze Pond, London SE1 9RT, UK.

[3] Department of Pathology, 8th Floor, Annexe Building, Tata Memorial Centre, Mumbai, India.

[4] Priority Research Centre for Cancer, School of Biomedical Sciences and Pharmacy, Faculty of Health, University of Newcastle, NSW, 2308, Australia.

[5] Peter Gorer Department of Immunobiology, King's College London, 2nd Floor, Borough Wing, Guy's Hospital, London, SE1 9RT, UK.

[6] Breast Cancer Now Research Unit, Innovation Hub, Cancer Centre at Guy's Hospital, King's College London, Faculty of Life Sciences and Medicine, London SE1 9RT, UK.

[$] corresponding author: anita.grigoriadis@kcl.ac.uk

Anita Grigoriadis

Cancer Bioinformatics, King's College London, Innovation Hub, Cancer Centre at Guy's Hospital, Great Maze Pond, London SE1 9RT, UK.

Email addresses:

GC: Gaurav.chatterjee@kcl.ac.uk

TP: Trupti.pai777@gmail.com

TH: Thomas.hardiman@kcl.ac.uk

KA: Kelly AveryKiejda@newcastle.edu.au

RS: rodney.scott@newcastle.edu.au

JS: jo.spencer@kcl.ac.uk

SP: Sarah.pinder@kcl.ac.uk

AG: Anita.grigoriadis@kcl.ac.uk

ABSTRACT (300/350 words)

Lymph node (LN) metastasis is an important prognostic parameter in breast carcinoma, a crucial site for tumour-immune cell interaction and a gateway for further dissemination of tumour cells to other metastatic sites. To gain insight into the underlying molecular changes from the pre-metastatic, via initial colonisation to the fully involved LN, we reviewed transcriptional research along the evolving microenvironment of LNs in human breast cancers patients. Gene expression studies were compiled and subjected to pathway-based analyses, with an emphasis on immune cell related genes. Of 366 studies, 14 performed genomewide gene expression comparisons and were divided into six clinical-biological scenarios capturing different stages of the metastatic pathway in the LN, as follows: metastatically involved LNs are compared to their patient-matched primary breast carcinomas (scenario 1) or the normal breast tissue (scenario 2). In scenario 3, uninvolved LNs were compared between LN-positive patients and LN-negative patients. Scenario 4 homed into the residual uninvolved portion of involved LNs and compared it to the patient-matched uninvolved LNs. Scenario 5 contrasted uninvolved and involved LNs; whilst in scenario 6 involved (sentinel) LNs were assessed between patients with other either positive or negative LNs (non-sentinel).

Gene lists from these chronological steps of LN metastasis indicated that gene patterns reflecting deficiencies in dendritic cells, hyper-proliferation of B cells parallel to tumour promoting pathways, including cell adhesion, extracellular matrix remodelling, cell motility and DNA repair play key-roles in the changing microenvironment of a pro-metastatic to a metastatically involved LN. Similarities between uninvolved LNs and the residual uninvolved portion of involved LNs hinted that LN alterations expose systemic tumour-related immune responses in breast cancer patients. Despite the diverse settings, gene expression patterns at different staged of metastatic colonisation in LNs were recognised, and may provide potential avenues for clinical interventions to counteract disease progression for breast cancer patients.

Keywords: expression, lymph node, premetastatic niche, breast cancer

INTRODUCTION

The lymph nodes (LNs) are functional units of the immune system that act as immunological hubs supporting the complex interactions between T cells, B cells, antigen presenting cells and stromal cells. LNs receive cells and potential immunogenic substances via the afferent lymphatics that drain the tissues and enter the LNs at the peripheral subcapsular sinus and also via the high endothelial venules which support lymphocyte entry from the blood (1, 2). The LN is a dynamic organ capable of undergoing dramatic remodelling, both in terms of architecture and function in response to pathological conditions such as inflammation or cancer (3). Many solid cancers spread through the lymphatic system to distant organs with the LNs typically serving as a first site of seeding outside primary tumour (4-6). For these tumours, the presence and extent of LN metastasis are markers of aggressive phenotype, often having an inverse linear relationship with prognosis (7-9). In breast carcinoma patients, metastasis to LN is an important factor for staging the tumour and routine assessment for invasive breast carcinoma patients includes histopathological assessment of presence of metastasis, the number of involved LNs and the presence or absence of extra-nodal extension (10).

Although the LN is a functional organ for tumour-immune system interaction and acts as a read-out for the systemic immune response, molecular characteristics of LNs centred around mutational alterations and structural genome rearrangements, whereas transcriptional research has been limited in both human and pre-clinical models (11). Most studies have aimed to identify molecular signatures associated with good and bad prognosis in primary breast tumours (12-18). While the genomes of relapsed or secondary breast cancers have revealed that metastases and primary

tumour are clonally related, share several driver mutations, and often acquired additional novel variants that are not present in the primary lesion (19).

In the metastatic LN, a multitude of factors play important roles in tilting the balance between pro-metastatic immunosuppression and anti-tumoural immune response (20-22). Given the significant implication of LN metastasis for systemic cancer burden, surprisingly few emphasis has been given to elucidate the underlying molecular signals and cellular alterations of the evolving LN microenvironment between the uninvolved (cancer-free) and the involved (metastatic) LNs in breast cancer patients. Some of these changes include lymphangiogenesis and increased lymph flow (23), recruitment and expansion of immunosuppressive cells (including myeloid-derived suppressor cells and regulatory T cells) (24), upregulation of chemokines and cytokines, blood vessel remodelling (25, 26) and a lower percentage of effector T cells (27). We recently comprehensively histologically characterised diverse immune and stromal features in primary tumours and their associated involved and uninvolved axillary LNs in a cohort of 309 invasive breast cancer patients (143/309 LN positive) (28), and observed that architectural alterations of the uninvolved LN are significant predictors for distant metastases. A similar finding of prognostic information from examination of the LN architecture was observed in melanoma (29). In preclinical mouse models, the involvement of innate lymphoid RORγt^{+} ILC3 cell, fibroblast reticular cells and cancer-associated fibroblasts in the induction of an immunosuppressive and pro-metastatic microenvironment in tumour-draining LNs was reported (30-32), while uninvolved regional LNs in rats with prostate tumours displayed varying degree of genetic changes depending on prostate tumour groups and their metastatic capacity (33).

With regards to emerging immunotherapy approaches, the LN microenvironment and the nature of the immune response have been identified as potent indicators of response to therapeutic interventions (34, 35). With the central position of the LN as an immune organ and as a gateway for further dissemination of tumour cells to other metastatic sites, we conducted a comprehensive review of existing gene expression-based research performed on LNs in human breast cancers. We categorised these gene expression studies along the evolving microenvironment of axillary metastases. By starting with early colonisation to the replacement of the entire LN with metastasis, these expression patterns capture information on the molecular mechanisms and changes in immune composition that allow the exploration of LNs as a pro-metastatic niche. Since the risk of developing distant metastasis and thus overall survival of patients with loco-regional metastatic breast cancer is typically poor, it is particularly important to establish whether transcriptomic patterns of metastasis might translate into new therapeutic strategies, including the successful implementation of immunotherapy.

MATERIAL AND METHODS

Literature Search and Data Collection

A review of the English literature was performed, focusing on gene expression data derived from human LN tissue and the primary lesion in breast cancer patients (if matched LN tissue was interrogated), using the combination of the following keywords: “breast cancer”, “metastasis”, “lymph nodes” and “gene expression” in “all fields” in PUBMED and Ovid MEDLINE ® (accessed on 13th October 2017 and revised on 5th

June 2018). All abstracts were manually screened and their methodologies were reviewed. Papers were selected if genomewide (i.e. microarray or RNA-sequencing based) gene expression analyses of LNs of breast cancer patients were performed (n=14). Studies of primary breast tumours and distant metastatic sites, which reported only the LN-status of the patients, were excluded (see consort diagram in Figure 1). The review was conducted according to the preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement (36).

Data analysis

Of total 366 papers screened, 14 studies were included in the review: Calvo *et al*. (37), Feng *et al*. (38), Hao *et al*. (39), Lähdesmäki *et al.* (40), Weigelt *et al*. (41), Ellsworth *et al*. (42), Vecchi *et al*. (43), Suzuki *et al*. (44), Mathe *et al*. (45), Zuckerman *et al*. (46), Blackburn *et al*. (47), Valente *et al*. (48), Rizwan *et al*. (49) (all of which performed microarray-based gene expression analyses) and Liang *et al*. (50), (which used 18-27 million paired-end riboZero RNA-sequencing). Genes with differential expression between the respective scenarios were obtained directly from the publications; no cut-offs were applied (Table 1). Using the biomaRt R package (51, 52) either gene names or microarray features were converted to ENSEMBL ID (ENSEMBL GRCh37.p13) (53) (Supplementary table 1,2,3). If microarray features could not be mapped (assuming that their sequences retired), they were excluded from further studies. Once an ENSEMBL ID list was created, HGNC symbols, genomic location and their common gene ontology terms were recorded. From these ENSEMBL gene lists, pathway analyses were conducted on de-regulated genes using the WebGestalt tool (54) (Supplementary table 4). The overrepresentation analysis (ORA) was applied based on the *Homo sapiens* Gene Ontology (GO) biological processes database. The whole

genome was used as a reference; all GO terms <0.05 FDR were extracted. To remove redundant GO terms, the Revigo tool, with parameter “small” was used (55). The resultant GO terms and differentially expressed genes were compared between the groups. To capture genes representative for specific immune cell populations, the gene-lists compounded from the studies were cross-referenced with published immune metagenes (56).

RESULTS AND DISCUSSION

Overview of expression profiling studies on LNs in Breast Cancer

A total of 14 genome-wide transcriptomic studies on LN samples were selected, to decipher the molecular features of the evolving LN microenvironment as a locoregional metastatic site (37-50). Each article published lists of genes specifically transcriptionally activated or repressed in LNs, ranging from cancer-free to metastatic settings. The cohorts were of mixed-receptor [Estrogen (ER), Progesterone (PR) and Human epidermal growth factor receptor (HER2)] invasive breast carcinomas, including two studies of invasive carcinomas of ductal/no special type only, and one exclusively examining triple negative breast carcinomas (TNBC). To paint a chronological picture of the changing microenvironment of the evolving metastatic LN, the studies were grouped into six “scenarios”, described below in detail (Table 1, Figure 2).

Scenario 1: Comparison between involved LN and primary breast carcinoma, the drivers of metastasis:

With the common aims of searching for drivers of metastatic progression, developing metastatic signatures predictive of distant metastasis (38, 43) and identifying molecular targets for metastasis-specific therapy or markers of resistance, eight of 14 studies captured transcriptional alterations between involved LNs and their patient-matched primary carcinoma. Expression patterns and gene regulatory pathways potentially driving metastatic dissemination were determined, while the point of acquiring metastatic efficiency in a primary tumour's timeline was intended to be revealed. These studies focussed on the cancerous tissue itself rather than the LN microenvironment; thus, the material selected for analyses had at least 70% tumour tissue, or laser microdissection was performed.

Although, high transcriptomic similarity between primary carcinoma and its corresponding LN metastasis were consistently observed (37, 40-42, 44), genes exclusively expressed in either of these two cancerous tissues was reported. Taking into consideration the clinical characteristics diversity of these cohorts, we asked whether any commonalities among activated or repressed genes could be established, potentially pointing collectively to deregulated biological themes. Among the eight studies, a total of 88 genes were found to be differentially expressed between the involved LN and the primary tumour in at least two studies, while the downregulation of 21 genes associated with cell- extracellular matrix (ECM) interaction, ECM remodelling, epithelial-mesenchymal transition (EMT) and loss of basement membrane function (57, 58) was common to 3 studies was (Supplementary table 1 & 2).

Downregulation of EMT-associated genes in the involved LN might suggest that, as the metastases becomes established, reversal of EMT and restoration to epithelial phenotype are essential for the successful colonisation (48). Stromal cells play a significant part in this process, particularly matrix metalloproteinases *MMP2* and *MMP7*, as these proteins are associated with the breakdown of the ECM, as well as innate immune response (59). CD10, a membrane metalloendopeptidase, is present at various stages of B cell maturation and of particular importance in LNs, where it is strongly expressed by germinal centre B cells, the most highly proliferative lymphocyte subset in LNs (60). CD10 was less abundant in involved LNs compared to the primary lesions in three studies (37, 42, 43), potentially pointing to a lack of differentiation potential of B cells.

Three genes, namely collagenase 11A1 (*COL11A1),* Asporin (*ASPN*) and Periostin (*POSTN*) were reported in 4 studies as having lower abundance in involved LNs compared to primary tumour tissue (38, 39, 42-44). All three genes function in remodelling ECM and ECM-associated protein degradation of the basement membrane. ECM remodelling is a well-established mechanistic prerequisite for dissemination of the primary cancer and genes involved in ECM are frequently part of metastatic gene sets in several other solid tumours (61). *COL11A1* promotes cell proliferation, migration and tumorigenesis of many human malignancies (62). This gene is currently investigated as a diagnostic marker for non-small cell lung carcinoma (NSCLC) and, by targeting *COL11A1,* chemoresistance might be overruled (63). The stromal expression of *ASPN* and *POSTN* has been shown to be associated with

aggressive tumour phenotypes and associated with poor prognosis in prostate and colorectal cancers, respectively (64, 65). If their lack of expression in involved LNs provides additive risk-information for disease progression warrants further investigation.

Complement component 7 (*C7*), a protein involved in the innate immune system, and part of the membrane attack complex that mediates lysis of pathogens, was the only gene of higher abundance in involved LNs reported in four studies (38, 42-44). Since *C7* may be related to processing and responding to different tumour neo-antigens present in involved LNs, its presence might reflect attempts of the involved LN to counterattack the metastatic colonisation.

Besides the malignant epithelial component, the transcriptional profiles of involved LNs almost always still harbour significant signals of immune and stroma cells. Among all eight studies, a total of 64 immune cell related genes was identified (Figure 3A & 3B, supplementary table 5), including upregulation of chemokines, ligands and receptors, cytotoxic CD8+ T cells, both immature and activated B cells, T cell receptor (TCR)- activation, MHC-class II, TH1 and TH2 in involved LNs. Conversely, genes down-regulated in involved LNs were associated with dendritic cells, mast cells and monocytes. Dendritic cells are antigen presenting cells that enter the LNs via the afferent lymphatics and that prime the effector T cells to initiate adaptive immune responses. Germinal centre responses are dependent on T cells activated by dendritic cells. A depletion of dendritic cells could represent a major immune escape mechanism in cancers (66), due to the lymphangiogenic responses in the metastatic

node (67). Thus, a dynamic interplay with the modulation of humoral and cellular immune responses, histologically corroborated by the reactive nodal changes with follicular, paracortical and sinusoidal hyperplasia is present in these involved LNs (28).

Overall our unifying analyses repeatedly demonstrated a consistent plasticity in ECM and immune cells in the metastatic LN tissue, despite the underlying molecular similarities between the primary carcinomas and patient-matched involved LNs. Cancer genomes reflect on clonal persistence and clonal extinction during cancer evolution (19). A recent comprehensive single cell analyses of chemoresistant TNBC supported an evolutionary model, in which an adaptive selection in the cancer genome is paralleled by an acquired transcriptional program, including ECM degradation and EMT (68). Given the remarkable molecular similarities between primary lesions and involved LNs, the metastatic genetic programme may be activated at an early stage during breast cancer development (15, 69), some cancerous cells may acquire their metastatic proficiency late due to clonal evolution (70), and as a sum are continually reshaping the metastatic molecular expression profiles (44). In parallel, as the metastatic potential of these cells evolves and increases over time, and the local microenvironment, through the interaction with endothelial, stromal and immune cells, carry significant determinants for successful colonisation in the LN.

Scenario 2: Comparison with normal breast tissue, pinpointing the changes in metastasis:

To decipher the remarkable similarities between a breast primary tumour and its LN metastasis, Mathe and colleagues (45) made multiple comparisons between normal breast tissue, LN-positive primary tumours, LN-negative primary tumours and LN metastases. Their hypothesis for identifying genes crucial for metastatic spread relied on: (i) genes differently expressed between primary tumour versus normal, tumour-adjacent breast tissue (NAT) in a LN-positive patient, followed by (ii) genes expressed in involved LN compared to normal breast tissue; and then (iii) selecting only those genes which were absent in primary tumours versus normal breast tissue in LN-negative patients. Through this step-wise approach, 14 genes were found commonly as downregulated in involved LNs (*APOD, MME, OMD, F2RL2, DCN, PTN, SFRP2, FMO1, OGN, SRPX, SPARCL1, MMP16, LRRC1, HMCN1*) (Supplementary table 3). *SPRX, SPARCL1, MMP16* and *HMCN1* are again involved in cell adhesion, ECM breakdown and organisation. *DCN* influences regulatory T cells (Treg) mediated immunosuppression, while *CD10*, as noted above, is essential for highly proliferative and pro-apoptotic germinal centre B cells (60, 71).

Performing an overrepresentation analysis using the GO database (54), pathways frequently deregulated in involved LNs in both scenarios 1 and 2 included ossification, cell adhesion, ECM organisation, cell proliferation, cell motility, apoptotic process and development of vasculature. Remodelling of the ECM and vascular proliferation are corroborated by the histological alterations in stromal architecture seen in LNs when metastasis manifests itself (Supplementary table 4) and have previously been linked to metastasis in multiple solid tumours (72).

In parallel, a delicate balance between helper and regulatory T cells seems to create a pro-metastatic immunosuppressive niche in the LN, as identified by 7 downregulated (*EGR1, RBMS3, CD34, IGF1, MEIS2, CMA1, DLC1)* and 5 upregulated *(MAD2L1, STAT1, KIF11, ANLN, DLGAP5**)*** genes associated with specific immune cell populations, especially T cell function including helper (*RBMS3, DLC1*) activated (*MAD2L1, KIF11, ANLN, DLGAP5*) and regulatory T cells. Different subsets of helper T cells, including Th17 and the heterogeneity of Tregs, are critical for cancer progression and metastasis (73, 74), again emphasising that the balance between different subset of helper and regulatory T cells is a crucial factor in successful colonisation.

Gene expression patterns across different phenotypical LNs groups:

By exclusively studying the involved LNs, key questions of: “when” does the LN microenvironment develop signals to potentially attract cancer cells, and when, why and how these cancer cells can home in such an immune cell-dominant environment, are omitted. LNs at different stages of colonisation provide the opportunity to obtain an insight in the underlying biology of evolving pre-metastatic setting. The following 4 scenarios adopted the diverse approaches across nodes of different status (Figure 2):

Scenario 3: By **comparing uninvolved LNs in LN-positive and LN-negative breast cancer patients**, the premetastatic niche and early genetic aberrations were interrogated for changes in immune response, vasculature, and cellular proliferation that were potentially measureable even before detectable metastasis had occurred. Here, molecular changes specific for a node-to-node manner and alterations systemically affect the regional nodes can be determined (46-48).

Scenario 4: **Comparison between the uninvolved, residual portion of a LN bearing metastatic carcinoma with patient matched negative nodes**, allowed identification of late-stage alterations in the secondary microenvironment, which may indirectly support metastatic growth (46, 48).

Scenario 5: By **comparing involved LNs with uninvolved LNs**, the alterations of immune and stromal cells within similar secondary microenvironment are captured (49).

Scenario 6: By **relating positive sentinel LNs in patients with additional, non-sentinel, positive LNs to patients with additional, non-sentinel, negative LNs**, genes patterns with increased risk of developing metastasis in other lymph nodes might be delineated (50).

Scenario 3: The uninvolved LN, the first step towards metastasis:

The first step in the colonisation of the LN by tumour cells is potentially the preparation of the LN microenvironment, even before the tumour cells arrive. Blackburn *et al*. (47) and Valente *et al.* (48), investigated the transcriptomic profiles of uninvolved LNs in LN-positive and LN-negative patients to identify early preparatory changes in the LN microenvironment. Both studies did not observe significant differences in gene expression patterns between the uninvolved LNs of LN-positive versus uninvolved LNs of LN-negative breast cancer patients(47, 48), and led the authors conclude that: (a) the physical presence of metastatic tumour cells may be crucial to elicit a pro-metastatic niche in the LNs; and (b) these pro-metastatic changes occur in a LN-to-LN manner and are not reflected systematically in uninvolved LNs in an otherwise LN-positive patient.

Studying the early metastatic changes, Zuckerman *et al*., followed a different approach by purifying immune cells from uninvolved sentinel and non-sentinel LNs. In uninvolved LNs (of entirely LN-negative patients), gene patterns were associated with immune cell regulation and signalling pathways such as antigen presentation (*HLA-DQA, HLA-A, HLA-DRB3*), lymphocyte activation (*HLA-DOA, IL23A, IL4, PLCG2, TICAM1*), cytokine-cytokine receptor interaction (*IL12RB2, IL4, CCR8, TNFRSF21, IL23A, IL3RA*) and pro-inflammatory TREM1 and IL-17 signalling (75, 76), indicating an effective antigen processing and anti-tumour response. TREM1 signalling activates monocyte-macrophage and neutrophil mediated immune response. IL-17 pathway stimulates Th17 cells to respond to a variety of foreign antigens and is involved in autoimmune diseases (77). An activation of such pro-inflammatory immune pathways in a LN-negative patient's LNs may facilitate an effective tumour response that prevents successful further spreading and colonisation of metastatic cells. In this context, breast cancer cells have been shown to hinder the functioning of dendritic cells and other antigen-processing cells (78). In contrast, the uninvolved LNs of LN-positive patients had higher levels of genes involved in relaxin signalling, which attracts mononuclear cells to create an immunosuppressive environment (79). The lack of effective immune responses, including antigen presentation, together with tumour promoting factors may all synergise to establish the necessary immunosuppressive pre-metastatic niche in the uninvolved LN of LN-positive patients. These molecular alterations may cause various architectural changes, including changes in size and location of germinal centres in uninvolved LNs of LN-positive breast cancer patients, as we have observed them (28).

Scenario 4: Residual portion of an involved LN, a surviving immune microenvironment:

A reflection of the vanishing immune cell microenvironment from the uninvolved to the involved LN, is provided by assessment of the residual portion of a LN where some colonisation by tumour cells has started (Figure 2 & 4). The uninvolved, 'normal' residual portion of an otherwise involved LN offers a unique snapshot of direct interaction between LN stromal and immune cells with tumour cells. To study the gene expression exclusively from this area of the LN, Valente *et al.* (48) confirmed the absence of tumour cells examination with AE1/AE3 immunohistochemical staining and laser microdissected the cancer-free tissue for RNA extraction. Similarly, Zuckerman and colleagues carefully selected, with flow cytometry-based sorting, only immune cells from residual LN materials (46). Most genes downregulated in the residual parts of involved LNs, when compared to completely uninvolved LNs, were involved in regulation of immune response (*HPGDS, STAB2, CLEC4M, PROS1, TFPI*), advocating a pro-metastatic immunosuppressive microenvironment. *STAB2*, a scavenger receptor, is known to regulate leukocyte trafficking in LNs through lymphatic endothelial cells (80), theoretically maintaining defence and tissue homeostasis, and in parallel spreading neoplastic cells. Similarly, in uninvolved LNs of otherwise LN-positive patients, pathways downregulated in the residual portion of positive LNs were pro-inflammatory immune related pathways like TREM1 signalling (*NOD2, TLR5)*, whilst the upregulated pathways were associated with cell cycle (*RAD51, KIF23, PLK4*), DNA repair (*RFC2, BRIP1*) and tumour promoting angiopoietin signalling (*RASA1, BRIP1*). In residual LN tissue (from nodes with metastatic tumour) compared to uninvolved LNs, B cell related genes (*AICDA, IGKC, IGKV1-5, IGKV3-20*), many of them specifically expressed in germinal centres, were

highly active. B cells and ectopic germinal centres have previously been linked to chronic inflammation and tumour promotion (81, 82), and may represent prognostic indicators for developing distant metastases (Figure 2 & 4) (28, 29). The upregulation of cell cycle and DNA repair pathways genes can further be linked to germinal centres, as these are zones of high proliferation. One might hypothesise, that in uninvolved LNs of LN-positive patients and in the residual 'normal' part of an involved LN, the upregulation of germinal centre B cell genes, in parallel to the dampening of antigen presentation and T cell priming, results in an altered tumour-promoting response, primarily mediated by B cells. Defective immune regulation in which B cell proliferation or humoral response is activated, in spite of the dampening of the antigen presentation and leukocyte activation, through some alternate pathways, could create a pro-metastatic environment. Furthermore, the abundance of kappa light chain genes as overexpressed in residual LN tissue point to an alternative B cell activation pathways biased towards B cells expressing kappa light chains and of oligoclonal nature. In the presence of B cell proliferation, it is essential to study markers such as PD-1, a negative regulator of B cell differentiation and expressed by the majority of T cells in germinal centres. B cells can both positively and negatively regulate T cell mediated antitumor immune responses, however their function in generating a specific pre-metastatic niche has yet to be established (67).

Scenario 5: From an uninvolved to an involved LN status:

To study the penultimate step in the evolving LN microenvironment one can look at the extreme endpoints, i.e. to capture transcriptional changes in the involved LN as a whole and compare with the uninvolved LN. Rizwan and colleagues mainly focussed

on change in collagen density in LNs in a murine metastatic breast cancer model, and examined expression patterns derived from publicly available microarray-based data (GSE4408), in which 16 involved and 3 uninvolved human LNs from breast cancer patients were compared (49). Ten of the fourteen genes transcriptionally activated in involved LNs, were fibronectin (*FN1),* three collagen genes *(COL1A2, COL1A1, COL3A1)* and 6 integrin family members (*ITGB5, ITGA2, ITGA9, ITGB7, ITGA2B, ITGA4*). All are key players in cell adhesion, cell-ECM interaction and ECM modulation (Supplementary table 3 & 4). Involved LNs displayed increased collagen I and basement membrane density in this murine metastatic breast cancer model. Increased collagen can promote tumour spread, not only by augmenting cell motility and regulating tumour promoting cell-ECM interactions, but also by altering immune responses, including switching the phenotypes of macrophages to a tumour-promoting M2 type (83) as well as a reduction of B cell follicles (49).

Scenario 6: The final step, can involved LNs send signals to other uninvolved LNs to promote tumour dissemination?

The number of involved LNs in breast cancer is associated with the risk of developing distant metastasis (7). The prediction of the extent and number of involved non-sentinel LNs by assessing the sentinel LN(s) would potentially have practical clinical importance, as axillary LN dissection in a group of sentinel LN-positive patients could be avoided (84, 85). The study by Liang and colleagues, although performed only on 6 patients, addressed the question of whether completely replaced LNs, especially the sentinel LNs, could send ‘signals’ to uninvolved LNs in preparation to disseminate the tumour cells (50). By comparing involved sentinel LNs in patients with additional

metastasis in non-sentinel LNs to those with otherwise negative axillary (non-sentinel) LNs, tumour promoting pathways were represented in non-sentinel LN-positive group, indicated by the expression of kallikrein subfamily members (*KLK10, KLK11, KLK12*, *KLK13*), proteolysis and steroid receptor signalling. In contrast, genes involved in plasma membrane and B cell receptor signalling, including *CD22, CD72, Igα, Igβ, CD19* and *CD21*, were depleted in parallel to *SYK, LYN, BTK, PTPN6*. In the group of patients with additional positive LNs, specific gene fusions were noted, especially involving *IGLL5*, a surrogate light chain involved in B cell development (86). Using immune metagenes denoting specific immune cell populations (56), an overlap between immature and activated B cells (*FCRLA, FAM129C, CD22, PAX5*), helper T cells (*SIGLEC10*), MDSCs (*CEACAM8, FCER2*), mast cells (*CLC, SIGLEC14*) and regulatory T cells (*CD72, IL9R*) (Supplementary table 5) was observed. Taken together, a recurrent theme for further tumour cell spreading emerges in these gene expression patterns, pointing strongly to a key role of B cells and germinal centres in LNs.

LN, a read-out for the systemic immune response?

Being an early site of tumour dissemination, the LN hosts a variety of tumour-immune system interactions. The ultimate question remains whether certain patterns in LNs of breast cancer patients' mirrors changes in the systemic immune response to the tumour in the organism. Valente and Blackburn argued that the physical presence of cancer cells in the LN is crucial for the pre-metastatic niche development and that the changes are therefore not systemic (47, 48). However, much recent research, such as the presence of similar immune gene sets in the uninvolved LNs in LN-positive

patients and the residual tissue of involved LNs (46), in addition to peripheral blood and to some extent in the immune compartment of the primary tumour (46), identified changes most likely indicative of a systemic effect in LN-positive patients. In keeping with this hypothesis, work on systemic immune responses to effective immunotherapies in preclinical murine breast cancer models has proven experimentally that changes in the immune composition persist in primary tumours, regional LNs, peripheral blood, bone marrow and other lymphoid organs (35).

Limitations

Despite the scarcity of expression data from LN tissue of breast cancer patients, together these data expose snap-shots of the steps in the molecular transitions that occur, starting from the uninvolved LN in LN-positive patients, to uninvolved residual tissue of involved LNs, to fully involved LNs, and finally the pro-disseminating signals in involved LNs. Ideally, all these comparisons should be examined within an individual patient's samples, to exclude patient-to-patient heterogeneity. Genomewide studies of whole LN samples mask effects in this highly spatially organised immune organ. Using sophisticated imaging technologies or single cell -omics analyses to capture earliest stages of LN metastasis, i.e. when tumour cells enter through the afferent lymphatic vessels and colonise in the subcapsular sinus (87), would provide valuable biological and potentially clinical relevant information.

CONCLUSION

The prognostic relevance of changes in uninvolved LNs is tantalising as it highlights the need of studying the interconnected roles of immune, stromal and endothelial cells

within this small immune organ as well as the whole immune system (28, 29). With the recent findings of the systemic orchestration of immune cells with effective immunotherapy (35), examination of local plus systemic tumour-immune cell interactions might hold the key for successful immunotherapeutic strategies. Although some patterns are evident from close scrutiny of existing literature, the 'premetastatic' LN represents an unmet knowledge gap; comprehensive cellular and molecular studies focusing on changes in different immune cell compartments at different time-points during the development of metastasis are needed to unlock this complicated biological process, both from a mechanistic and therapeutic point of view.

DECLARATIONS

Ethical Approval and Consent to Participate

Not applicable

Consent for publication

Not applicable

Availability of data and material

Not applicable

Competing interests

The authors declare no competing interests.

Acknowledgement and Funding

AG is supported by Breast Cancer Now Unit funding at King's College London and Cancer Research UK King's Health Partners Centre at King's College London. Thomas Hardiman is funded by the MRC-DTP. GC and TP have been recipients of scholarship from Tata Memorial Centre. We thank Patrycja Gazinska for providing the H&E images.

Statement of author contributions

Concept and design: GC, TP, AG. Data acquisition: GC, TG, TH, AM, RJS, AG. Data analysis and interpretation: GC, TP, TH, JS, SP, AG. Wrote the manuscript: GC, TP, SP, AG. Critically reviewed the manuscript: GC, TP, JS, SP, AG.

REFERENCES:

1. Ji R-C. Lymph Nodes and Cancer Metastasis: New Perspectives on the Role of Intranodal Lymphatic Sinuses. International Journal of Molecular Sciences. 2017;18(1):51.
2. Ruddle NH. Lymphatic vessels and tertiary lymphoid organs. The Journal of clinical investigation. 2014;124(3):953-9.
3. Zhu M, Fu YX. The role of core TNF/LIGHT family members in lymph node homeostasis and remodeling. Immunological reviews. 2011;244(1):75-84.
4. Lund AW, Wagner M, Fankhauser M, Steinskog ES, Broggi MA, Spranger S, et al. Lymphatic vessels regulate immune microenvironments in human and murine melanoma. The Journal of clinical investigation. 2016;126(9):3389-402.
5. Tauchi Y, Tanaka H, Kumamoto K, Tokumoto M, Sakimura C, Sakurai K, et al. Tumor-associated macrophages induce capillary morphogenesis of lymphatic endothelial cells derived from human gastric cancer. Cancer science. 2016;107(8):1101-9.
6. Ji RC. Lymph node lymphangiogenesis: a new concept for modulating tumor metastasis and inflammatory process. Histology and histopathology. 2009;24(3):377-84.
7. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, et al. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more “personalized” approach to cancer staging. CA: a cancer journal for clinicians. 2017;67(2):93-9.
8. Dickson PV, Gershenwald JE. Staging and Prognosis of Cutaneous Melanoma. Surgical oncology clinics of North America. 2011;20(1):1-17.
9. Carter CL, Allen C, Henson DE. Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases. Cancer. 1989;63(1):181-7.
10. The Royal College of Pathologists. [Available from: https://www.rcpath.org/resourceLibrary/g148-breastdataset-hires-jun16-pdf.html]
11. Kroigard AB, Larsen MJ, Thomassen M, Kruse TA. Molecular Concordance Between Primary Breast Cancer and Matched Metastases. The breast journal. 2016;22(4):420-30.

12. Ellsworth RE, Field LA, Love B, Kane JL, Hooke JA, Shriver CD. Differential Gene Expression in Primary Breast Tumors Associated with Lymph Node Metastasis. International journal of breast cancer. 2011;2011:142763.

13. Zhang Z, Yamashita H, Toyama T, Sugiura H, Ando Y, Mita K, et al. ATBF1-a messenger RNA expression is correlated with better prognosis in breast cancer. Clinical cancer research : an official journal of the American Association for Cancer Research. 2005;11(1):193-8.

14. Sotiriou C, Neo SY, McShane LM, Korn EL, Long PM, Jazaeri A, et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proceedings of the National Academy of Sciences of the United States of America. 2003;100(18):10393-8.

15. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, et al. A gene-expression signature as a predictor of survival in breast cancer. The New England journal of medicine. 2002;347(25):1999-2009.

16. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature. 2002;415(6871):530-6.

17. Sotiriou C, Pusztai L. Gene-Expression Signatures in Breast Cancer. New England Journal of Medicine. 2009;360(8):790-800.

18. Muthukaruppan A, Lasham A, Blenkiron C, Woad KJ, Black MA, Knowlton N, et al. Gene expression profiling of breast tumours from New Zealand patients. The New Zealand medical journal. 2017;130(1464):40-56.

19. Yates LR, Knappskog S, Wedge D, Farmery JHR, Gonzalez S, Martincorena I, et al. Genomic Evolution of Breast Cancer Metastasis and Relapse. Cancer Cell. 2017;32(2):169-84.e7.

20. Jones D, Pereira ER, Padera TP. Growth and Immune Evasion of Lymph Node Metastasis. Frontiers in Oncology. 2018;8(36).

21. Liu Y, Cao X. Characteristics and Significance of the Pre-metastatic Niche. Cancer Cell. 2016;30(5):668-81.

22. Sleeman JP. The lymph node pre-metastatic niche. Journal of molecular medicine (Berlin, Germany). 2015;93(11):1173-84.

23. Harrell MI, Iritani BM, Ruddell A. Tumor-induced sentinel lymph node lymphangiogenesis and increased lymph flow precede melanoma metastasis. The American journal of pathology. 2007;170(2):774-86.

24. Ogawa F, Amano H, Eshima K, Ito Y, Matsui Y, Hosono K, et al. Prostanoid induces premetastatic niche in regional lymph nodes. The Journal of clinical investigation. 2014;124(11):4882-94.

25. Chung MK, Do IG, Jung E, Son YI, Jeong HS, Baek CH. Lymphatic vessels and high endothelial venules are increased in the sentinel lymph nodes of patients with oral squamous cell carcinoma before the arrival of tumor cells. Annals of surgical oncology. 2012;19(5):1595-601.

26. Qian CN, Berghuis B, Tsarfaty G, Bruch M, Kort EJ, Ditlev J, et al. Preparing the "soil": the primary tumor induces vasculature reorganization in the sentinel lymph node before the arrival of metastatic cancer cells. Cancer research. 2006;66(21):10365-76.

27. Kohrt HE, Nouri N, Nowels K, Johnson D, Holmes S, Lee PP. Profile of immune cells in axillary lymph nodes predicts disease-free survival in breast cancer. PLoS medicine. 2005;2(9):e284.

28. Grigoriadis A, Gazinska P, Pai T, Irhsad S, Wu Y, Millis R, et al. Histological scoring of immune and stromal features in breast and axillary lymph nodes is prognostic for distant metastasis in lymph node-positive breast cancers. J Pathol Clin Res. 2018;4(1):39-54.

29. Abbott J, Buckley M, Taylor LA, Xu G, Karakousis G, Czerniecki BJ, et al. Histological Immune Response Patterns in Sentinel Lymph Nodes Involved by Metastatic Melanoma and Prognostic Significance. Journal of cutaneous pathology. 2018.

30. Riedel A, Shorthouse D, Haas L, Hall BA, Shields J. Tumor-induced stromal reprogramming drives lymph node transformation. Nature immunology. 2016;17(9):1118-27.

31. Costa A, Kieffer Y, Scholer-Dahirel A, Pelon F, Bourachot B, Cardon M, et al. Fibroblast Heterogeneity and Immunosuppressive Environment in Human Breast Cancer. Cancer Cell. 2018.

32. Irshad S, Flores-Borja F, Lawler K, Monypenny J, Evans R, Male V, et al. RORgammat(+) Innate Lymphoid Cells Promote Lymph Node Metastasis of Breast Cancers. Cancer Res. 2017;77(5):1083-96.

33. Strömvall K, Thysell E, Halin Bergström S, Bergh A. Aggressive rat prostate tumors reprogram the benign parts of the prostate and regional lymph nodes prior to metastasis. PLOS ONE. 2017;12(5):e0176679.

34. Chen J, Wang L, Yao Q, Ling R, Li K, Wang H. Drug concentrations in axillary lymph nodes after lymphatic chemotherapy on patients with breast cancer. Breast Cancer Research. 2004;6(4):R474.

35. Spitzer MH, Carmi Y, Reticker-Flynn NE, Kwek SS, Madhireddy D, Martins MM, et al. Systemic Immunity Is Required for Effective Cancer Immunotherapy. Cell. 2017;168(3):487-502.e15.

36. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Medicine. 2009;6(7):e1000097.

37. Calvo J, Sanchez-Cid L, Munoz M, Lozano JJ, Thomson TM, Fernandez PL. Infrequent loss of luminal differentiation in ductal breast cancer metastasis. PLoS One. 2013;8(10):e78097.

38. Feng Y, Sun B, Li X, Zhang L, Niu Y, Xiao C, et al. Differentially expressed genes between primary cancer and paired lymph node metastases predict clinical outcome of node-positive breast cancer patients. Breast cancer research and treatment. 2007;103(3):319-29.

39. Hao X, Sun B, Hu L, Lahdesmaki H, Dunmire V, Feng Y, et al. Differential gene and protein expression in primary breast malignancies and their lymph node metastases as revealed by combined cDNA microarray and tissue microarray analysis. Cancer. 2004;100(6):1110-22.

40. Lahdesmaki H, Hao X, Sun B, Hu L, Yli-Harja O, Shmulevich I, et al. Distinguishing key biological pathways between primary breast cancers and their lymph node metastases by gene function-based clustering analysis. International journal of oncology. 2004;24(6):1589-96.

41. Weigelt B, Wessels LF, Bosma AJ, Glas AM, Nuyten DS, He YD, et al. No common denominator for breast cancer lymph node metastasis. British journal of cancer. 2005;93(8):924-32.

42. Ellsworth RE, Seebach J, Field LA, Heckman C, Kane J, Hooke JA, et al. A gene expression signature that defines breast cancer metastases. Clinical & experimental metastasis. 2009;26(3):205-13.

43. Vecchi M, Confalonieri S, Nuciforo P, Vigano MA, Capra M, Bianchi M, et al. Breast cancer metastases are molecularly distinct from their primary tumors. Oncogene. 2008;27(15):2148-58.

44. Suzuki M, Tarin D. Gene expression profiling of human lymph node metastases and matched primary breast carcinomas: clinical implications. Molecular oncology. 2007;1(2):172-80.

45. Mathe A, Wong-Brown M, Morten B, Forbes JF, Braye SG, Avery-Kiejda KA, et al. Novel genes associated with lymph node metastasis in triple negative breast cancer. Scientific Reports. 2015;5:15832.

46. Zuckerman NS, Yu H, Simons DL, Bhattacharya N, Carcamo-Cavazos V, Yan N, et al. Altered local and systemic immune profiles underlie lymph node metastasis in breast cancer patients. International journal of cancer. 2013;132(11):2537-47.

47. Blackburn HL, Ellsworth DL, Shriver CD, Ellsworth RE. Breast Cancer Metastasis to the Axillary Lymph Nodes: Are Changes to the Lymph Node "Soil" Localized or Systemic? Breast cancer : basic and clinical research. 2017;11:1178223417691246.

48. Valente AL, Kane JL, Ellsworth DL, Shriver CD, Ellsworth RE. Molecular response of the axillary lymph node microenvironment to metastatic colonization. Clinical & experimental metastasis. 2014;31(5):565-72.

49. Rizwan A, Bulte C, Kalaichelvan A, Cheng M, Krishnamachary B, Bhujwalla ZM, et al. Metastatic breast cancer cells in lymph nodes increase nodal collagen density. Sci Rep. 2015;5:10002.

50. Liang F, Qu H, Lin Q, Yang Y, Ruan X, Zhang B, et al. Molecular biomarkers screened by next-generation RNA sequencing for non-sentinel lymph node status prediction in breast cancer patients with metastatic sentinel lymph nodes. World Journal of Surgical Oncology. 2015;13(1):258.

51. Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, et al. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics. 2005;21(16):3439-40.

52. Durinck S, Spellman PT, Birney E, Huber W. Mapping Identifiers for the Integration of Genomic Datasets with the R/Bioconductor package biomaRt. Nature protocols. 2009;4(8):1184-91.

53. Flicek P, Amode MR, Barrell D, Beal K, Brent S, Carvalho-Silva D, et al. Ensembl 2012. Nucleic acids research. 2012;40(Database issue):D84-D90.

54. Wang J, Vasaikar S, Shi Z, Greer M, Zhang B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. Nucleic acids research. 2017;45(W1):W130-W7.

55. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. PLOS ONE. 2011;6(7):e21800.

56. Angelova M, Charoentong P, Hackl H, Fischer ML, Snajder R, Krogsdam AM, et al. Characterization of the immunophenotypes and antigenomes of colorectal cancers reveals distinct tumor escape mechanisms and novel targets for immunotherapy. Genome biology. 2015;16(1):64.

57. Pickup MW, Mouw JK, Weaver VM. The extracellular matrix modulates the hallmarks of cancer. EMBO Reports. 2014;15(12):1243-53.

58. Holle AW, Young JL, Spatz JP. In vitro cancer cell–ECM interactions inform in vivo cancer treatment. Advanced Drug Delivery Reviews. 2016;97:270-9.

59. Edman K, Furber M, Hemsley P, Johansson C, Pairaudeau G, Petersen J, et al. The Discovery of MMP7 Inhibitors Exploiting a Novel Selectivity Trigger. ChemMedChem. 2011;6(5):769-73.

60. Höller S, Horn H, Lohr A, Mäder U, Katzenberger T, Kalla J, et al. A cytomorphological and immunohistochemical profile of aggressive B-cell lymphoma: high clinical impact of a cumulative immunohistochemical outcome predictor score. Journal of Hematopathology. 2009;2(4):187-94.

61. Kondoh N, Ishikawa T, Ohkura S, Arai M, Hada A, Yamazaki Y, et al. Gene expression signatures that classify the mode of invasion of primary oral squamous cell carcinomas. Molecular carcinogenesis. 2008;47(10):744-56.

62. Li A, Li J, Lin J, Zhuo W, Si J. COL11A1 is overexpressed in gastric cancer tissues and regulates proliferation, migration and invasion of HGC-27 gastric cancer cells in vitro. Oncology reports. 2017;37(1):333-40.

63. Shen L, Yang M, Lin Q, Zhang Z, Zhu B, Miao C. COL11A1 is overexpressed in recurrent non-small cell lung cancer and promotes cell proliferation, migration, invasion and drug resistance. Oncology reports. 2016;36(2):877-85.

64. Rochette A, Boufaied N, Scarlata E, Hamel L, Brimo F, Whitaker HC, et al. Asporin is a stromally expressed marker associated with prostate cancer progression. British journal of cancer. 2017;116(6):775-84.

65. Oh HJ, Bae JM, Wen XY, Cho NY, Kim JH, Kang GH. Overexpression of POSTN in Tumor Stroma Is a Poor Prognostic Indicator of Colorectal Cancer. Journal of pathology and translational medicine. 2017;51(3):306-13.

66. Kusume A, Sasahira T, Luo Y, Isobe M, Nakagawa N, Tatsumoto N, et al. Suppression of dendritic cells by HMGB1 is associated with lymph node metastasis of human colon cancer. Pathobiology : journal of immunopathology, molecular and cellular biology. 2009;76(4):155-62.

67. Balsat C, Blacher S, Herfs M, Van de Velde M, Signolle N, Sauthier P, et al. A specific immune and lymphatic profile characterizes the pre-metastatic state of the sentinel lymph node in patients with early cervical cancer. Oncoimmunology. 2017;6(2):e1265718.

68. Kim C, Gao R, Sei E, Brandt R, Hartman J, Hatschek T, et al. Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. Cell. 2018;173(4):879-93.e13.

69. Ramaswamy S, Ross KN, Lander ES, Golub TR. A molecular signature of metastasis in primary solid tumors. Nature genetics. 2003;33(1):49-54.

70. Montel V, Huang TY, Mose E, Pestonjamasp K, Tarin D. Expression profiling of primary tumors and matched lymphatic and lung metastases in a xenogeneic breast cancer model. The American journal of pathology. 2005;166(5):1565-79.

71. Oh E, Choi IK, Hong J, Yun CO. Oncolytic adenovirus coexpressing interleukin-12 and decorin overcomes Treg-mediated immunosuppression inducing potent antitumor effects in a weakly immunogenic tumor model. Oncotarget. 2017;8(3):4730-46.

72. Cox TR, Erler JT. Remodeling and homeostasis of the extracellular matrix: implications for fibrotic diseases and cancer. Disease models & mechanisms. 2011;4(2):165-78.

73. Ward-Hartstonge KA, Kemp RA. Regulatory T-cell heterogeneity and the cancer immune response. Clin Trans Immunol. 2017;6:e154.

74. Marshall EA, Ng KW, Kung SH, Conway EM, Martinez VD, Halvorsen EC, et al. Emerging roles of T helper 17 and regulatory T cells in lung cancer progression and metastasis. Molecular cancer. 2016;15(1):67.

75. Gibot S, Cravoisy A, Levy B, Bene M-C, Faure G, Bollaert P-E. Soluble Triggering Receptor Expressed on Myeloid Cells and the Diagnosis of Pneumonia. New England Journal of Medicine. 2004;350(5):451-8.

76. Qian Y, Kang Z, Liu C, Li X. IL-17 signaling in host defense and inflammatory diseases. Cellular And Molecular Immunology. 2010;7:328.

77. Liu CJ, Tsai CY, Chiang SH, Tang SJ, Chen NJ, Mak TW, et al. Triggering receptor expressed on myeloid cells-1 (TREM-1) deficiency augments BAFF production to promote lupus progression. Journal of autoimmunity. 2017;78:92-100.

78. Tourkova IL, Shurin GV, Ferrone S, Shurin MR. Interferon regulatory factor 8 mediates tumor-induced inhibition of antigen processing and presentation by dendritic cells. Cancer immunology, immunotherapy : CII. 2009;58(4):567-74.

79. Figueiredo KA, Mui AL, Nelson CC, Cox ME. Relaxin stimulates leukocyte adhesion and migration through a relaxin receptor LGR7-dependent mechanism. The Journal of Biological Chemistry. 2006;281(6):3030-9.

80. Ji RC. Lymph Nodes and Cancer Metastasis: New Perspectives on the Role of Intranodal Lymphatic Sinuses. International journal of molecular sciences. 2016;18(1).

81. de Visser KE, Korets LV, Coussens LM. De novo carcinogenesis promoted by chronic inflammation is B lymphocyte dependent. Cancer Cell. 2005;7(5):411-23.

82. William J, Euler C, Christensen S, Shlomchik MJ. Evolution of autoantibody responses via somatic hypermutation outside of germinal centers. Science. 2002;297(5589):2066-70.

83. Fang M, Yuan J, Peng C, Li Y. Collagen as a double-edged sword in tumor progression. Tumour Biology. 2014;35(4):2871-82.

84. Hwang RF, Krishnamurthy S, Hunt KK, Mirza N, Ames FC, Feig B, et al. Clinicopathologic factors predicting involvement of nonsentinel axillary nodes in women with breast cancer. Annals of surgical oncology. 2003;10(3):248-54.

85. Van Zee KJ, Manasseh DM, Bevilacqua JL, Boolbol SK, Fey JV, Tan LK, et al. A nomogram for predicting the likelihood of additional nodal metastases in breast

cancer patients with a positive sentinel node biopsy. Annals of surgical oncology. 2003;10(10):1140-51.

86. Pridans C, Holmes ML, Polli M, Wettenhall JM, Dakic A, Corcoran LM, et al. Identification of Pax5 Target Genes in Early B Cell Differentiation. The Journal of Immunology. 2008;180(3):1719-28.

87. Pereira ER, Jones D, Jung K, Padera TP. The lymph node microenvironment and its role in the progression of metastatic cancer. Seminars in cell & developmental biology. 2015;38:98-105.

Table 1: Genomewide expression studies of LNs of breast carcinoma patients.

CLINICAL QUESTION	STUDY	BREAST CARCINOMA	SAMPLE COHORT	RESULTS
Scenario 1 Involved lymph node (ILN) versus primary tumours (PT)	Calvo *et al*. (37), 2013	IDC	18 PT vs matching ILN	Infrequent loss of luminal differentiation in metastatic LN
	Feng *et al*. (38), 2007	IDC	26 PT vs matching ILN	79 DEG
	Hao *et al*. & Lähdesmäki *et al*.(39, 40), 2004	Invasive BC	9 PT vs matching ILN	280 DEG
	Weigelt *et al* (41), 2005	Invasive BC	15 PT vs matching ILN	No classifier or single gene could discriminate
	Ellsworth *et al*. (42), 2009	Invasive BC	20 PT vs matching ILN	51 DEG
	Vecchi *et al*. (43), 2008	Invasive BC	26 PT vs matching ILN	270 DEG
	Suzuki *et al*. (44), 2007	Invasive BC	10 PT vs matching ILN	84 DEG
Scenario 2 Involved LN versus normal adjacent breast tissue (NAT)	Mathe *et al*. (45), 2015	TNBC	15 ILN vs 17 NAT	83 genes were significantly associated with LN metastasis
Scenario 3 Uninvolved LN in LN-positive versus LN-negative patients	Zuckerman *et al*. (46), 2013	Invasive BC	11 PT, 30 LN, 21 PB	116/ 219 DEG (SLN/ NSLN respectively)
	Blackburn *et al*. (47), 2017	Invasive BC	24 LN from NP vs 40 LN from NN	No genes were differentially expressed with stringent FDR
Scenario 4 Uninvolved residual portion of involved LN versus uninvolved LN	Valente *et al*. (48), 2014	Invasive BC	20 matched pairs of involved and uninvolved LN	22 DEG
	Zuckerman *et al*. (46), 2013	Invasive BC	11 PT, 30 LN, 21 PB	103 DEG
Scenario 5 Involved LN versus uninvolved LN	Rizwan *et al*. (49), 2015	Invasive BC	16 involved vs 3 uninvolved LN	13 DEG
Scenario 6 Positive sentinel LNs in patients with additional, non-sentinel, positive LNs to patients with additional, non-sentinel, negative LNs	Liang *et al*. (50), 2015	Invasive BC	3 NSLN+ SLN vs 3 NSLN- SLN	160 DEG

Figure legends

Figure 1: Systematic review flowchart in accordance to the PRISMA statement (36) for the gene expression studies performed on LNs in human breast cancer patients. Total 14 studies were included after the procedure of searching, screening and excluding from the English literature database. Thirteen of these studies were subjected to quantitative analysis.

Figure 2: Different scenarios studying lymph nodes, breast cancers and normal tissue. Six scenario depicted different comparisons (indicated by green arrows) = Scenario 1: involved lymph node versus primary tumour (# of studies= 8); Scenario 2: involved lymph node versus normal breast tissue (# of studies= 1); Scenario 3: uninvolved LNs in LN-positive patients versus uninvolved LNs in LN-negative patients (# of studies= 2); Scenario 4: uninvolved residual portion of involved LN versus patient-matched uninvolved LN (# of studies= 2); Scenario 5: involved LN versus patient-matched uninvolved LNs (# of studies= 1); Scenario 6: involved sentinel LNs in patients with additional, non-sentinel, positive LNs versus involved sentinel LNs in patients with additional, non-sentinel, negative LNs (# of studies= 1). Tumours are shown in orange, red and green denote involved and uninvolved LNs, respectively. In scenario 4, the shaded portion represents the uninvolved residual portion of an involved LN.

Figure 3: Immune cell composition in different scenarios. A: The percentage of genes representing specific immune cell populations in each of the scenarios is

shown. **B.** The proportion of different immune cell populations among all the immune-related genes in each scenario. (Scenario 4 was omitted, as reported 103 differentially expressed genes could not be retrieved from the original study).

Figure 4: Chronological steps of lymph node metastasis (H&E stain). (A) An uninvolved axillary LN with no evidence of tumour cells (0.7X). **(B)** Partial colonisation of a LN with significant amount of residual uninvolved LN tissue (black arrowhead) and two nodules of metastasis (black arrows) are depicted (0.5X). Inset shows tumour cells mixed with background immune cells (20X). **(C)** A lymph node with near total replacement of normal lymph nodal tissue (1X). The inset displays a higher power magnification of tumour cells (10X). All images were captured by Nanozoomer and viewed in NDP.view2 software (Hamamatsu).

Table 1: Genomewide expression studies of LNs of breast carcinoma patients. BC = Breast carcinoma, DEG = Differentially expressed genes, IDC = Invasive ductal carcinoma (no special type), ILN = Involved LN, LN = lymph node, NAT = Normal adjacent breast tissue, NN = Node negative patients, NP = Node positive patients, NSLN = Non-sentinel lymph node, PT = Primary tumour, SLN = Sentinel lymph node.

Supplementary information

Supplementary table 1: Gene list compiled from all the studies included in scenario 1 (Involved LN versus primary breast tumour).

Supplementary table 2: List of genes found to be differentially expressed in multiple studies included in scenario 1 (Involved LN versus primary breast tumour).

Supplementary table 3: Fully compiled gene list across all scenarios.

Supplementary table 4: Pathway-based analysis of the differentially expressed genes across all the scenarios.

Supplementary table 5: Differentially expressed genes representing specific immune cell populations across all the scenarios.